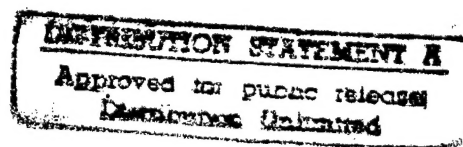# NATIONAL COMMUNICATIONS SYSTEM

TECHNICAL INFORMATION BULLETIN 93-3

ENHANCED FACSIMILE SERVICES

JANUARY 1993

OFFICE OF THE MANAGER
NATIONAL COMMUNICATIONS SYSTEM
701 SOUTH COURT HOUSE ROAD
ARLINGTON, VA 22204-2198

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE January 1993	3. REPORT TYPE AND DATES COVERED Final Report

4. TITLE AND SUBTITLE

Enhanced Facsimile Services

5. FUNDING NUMBERS

DCA100-91-C-0031

6. AUTHOR(S)

Stephen Perschau

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Delta Information Systems, Inc.
Horsham Business Center, Bldg 3, Ste 120
300 Welsh Road
Horsham, PA 19044-2273

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Communications System
Office of Technology and Standards Division
701 South Court House Road
Arlington, Virginia 22204-2198

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NCS TIB 93-3

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Facsimile apparatus will begin to have access to local area networks, and store-and-forward facsimile network servies. Plus, they will begin to have access to Digital Circuit Multiplying Equipment, Packetized Circuit Multiplying Equipment and Packet Assemblers/Disassemblers. This task investigates using Optical Character Recongnition to help interconnect facsimile to these services. This report discusses using OCR to convey instructions from a facsimile terminal to a store-and-forward; discusses how facsimiles, including graphics, might be converted to text documents; discusses how facsimile modems and enhanced services affect long distance communication equipment, and how the long distance equipment might determine the modem modulation method used. It also provides recommendations for using OCR and character or binary encodings for message transfers between facsimile terminals and store-and-forward systems. It also recommends how long distance equipments might better identify facsimile and data modem traffic.

Most store-and-forward terminals communicate indirectly in multipoint configurations. Messages between terminals can sometimes be stored in the network. These messages usually stay in the network until the recipient retrieves them.

14. SUBJECT TERMS

Optical Character Recognition (OCR)
Group 3 and Group 4
Facsimile

15. NUMBER OF PAGES
50 pages

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASS	UNCLASS	UNCLASS	UNLIMITED

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to *stay within the lines* to meet *optical scanning requirements*.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - Leave blank.
NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

ENHANCED FACSIMILE SERVICES

JANUARY 1993

PROJECT OFFICER

[signature: Stephen Perschau]

STEPHEN PERSCHAU
Computer Scientist
Office of Technology
 and Standards

APPROVED FOR PUBLICATION:

[signature: Dennis Bodson]

DENNIS BODSON
Assistant Manager
Office of Technology
 and Standards

FOREWORD

Among the responsibilities assigned to the Office of the Manager, National Communications System, is the management of the Federal Telecommunication Standards Program. Under this program, the NCS, with the assistance of the Federal Telecommunication Standards Committee identifies, develops, and coordinates proposed Federal Standards which either contribute to the interoperability of functionally similar Federal telecommunication systems or to the achievement of a compatible and efficient interface between computer and telecommunication systems. In developing and coordinating these standards, a considerable amount of effort is expended in initiating and pursuing joint standards development efforts with appropriate technical committees of the International Organization for Standardization, and the International Telegraph and Telephone Consultative Committee of the International Telecommunication Union. This Technical Information Bulletin presents and overview of an effort which is contributing to the development of compatible Federal, national, and international standards in the area of facsimile. It has been prepared to inform interested Federal activities of the progress of these efforts. Any comments, inputs or statements of requirements which could assist in the advancement of this work are welcome and should be addressed to:

Office of the Manager
National Communications System
Attn: NT
701 S. Court House Road
Arlington, VA 22204-2198

ENHANCED FACSIMILE SERVICES

January 1993

SUBMITTED TO:
NATIONAL COMMUNICATIONS SYSTEM
Office of Technology and Standards
ARLINGTON, VA 22204-2198

Contracting Agency:
DEFENSE INFORMATION SYSTEMS AGENCY
Contract Number - DCA100-91-C-0031
Task Number - 91-001-4

DELTA INFORMATION SYSTEMS, INC
Horsham Business Center, Bldg. 3, Ste 120
300 Welsh Road
Horsham, PA 19044-2273

TEL: (215) 657-5270   FAX: (215) 657-5273

CONTENTS

ILLUSTRATIONS

TABLES

1.0 INTRODUCTION

This document summarizes work performed by Delta Information Systems, Inc. for the Office of Technology and Standards of the National Communications System, an organization of the U. S. government. The effort was specified by task 1, subtask 4 of contract number DCA100-91-C-0031. With the development of new and enhanced facsimile services, facsimile is quickly advancing beyond the original point-to-point image transfer. Facsimile apparatus will begin to have access to local area networks, and store-and-forward facsimile network services. Plus, they will begin to have access to Digital Circuit Multiplying Equipment (DCME), Packetized Circuit Multiplying Equipment (PCME) and Packet Assemblers/Disassemblers (FPADS). This task investigates using Optical Character Recognition (OCR) to help interconnect facsimile to these services.

1.1 Report Organization

This report has seven sections:

1. Introduction
2. Optical Character Recognition
3. Instruction Stream Recognition
4. Converting Facsimiles to Text Documents
5. Comparison of Character and Binary Encoding Methods
6. Facsimile Traffic Compression over Long Distances
7. Summary and Recommendations

Section 1.0, "Introduction," provides background information and discusses this reports organization.

Section 2.0, "Optical Character Recognition," discusses OCR, and its speed and accuracy.

Section 3.0, "Instruction Stream Recognition," discusses using OCR to convey instructions from a facsimile terminal to a store-and-forward.

Section 4.0, "Converting Facsimiles to Text Documents," discusses how facsimiles, including graphics, might be converted to text documents.

Section 5.0, "Comparison of Character and Binary Encoding Methods," compares using character and binary encoding methods for conveying instruction streams.

Section 6.0, "Facsimile Traffic Compression over Long Distances", discusses how facsimile modems and enhanced services affect long distance communication equipment, and how the long distance equipment might determine the modem modulation method used.

Section 7.0, "Summary and Recommendations," summarizes the report and provides recommendations for using OCR and character or binary encodings for message transfers between facsimile terminals and store-and-forward systems. It also recommends how long distance equipments might better identify facsimile and data modem traffic.

1.2 Background

Most store-and-forward terminals communicate indirectly in multipoint configurations. Messages between terminals can sometimes be stored in the network. These messages usually stay in the network until the recipient retrieves them. Using facsimile equipment on store-and-forward systems poses several challenges. Most facsimile equipments (Group 3 and Group 4) communicate real-time using point-to-point configurations. On store-and-forward systems, real-time communications between sending and receiving equipments is sometimes impractical. Plus, establishing common capabilities among several facsimile terminals could be difficult in a multipoint environment, when messages are stored, or both. In addition, facsimile messages could be sent to character-only capable

terminals (e.g., teletypes). As a result, some facsimile imagery (photos, etc.) should probably be rendered using characters.

Several mechanisms could be used to overcome these difficulties, some short-term, some long-term. Short-term mechanisms would stress maximizing a service's functionality without requiring facsimile protocol modifications. Long-term mechanisms would allow protocol modifications and would provide an evolutionary path for additional capabilities. Both would stress compliance with existing standards.

1.2.1 Short Term Mechanisms

In one short-term mechanism, the facsimile equipments would believe that they are communicating with another facsimile equipment. In actuality, they would communicate with a service's User Agent (UA). The UA would be tailored to facsimile communication. The service's user would register with the UA to receive facsimile messages. The UA would assign the user an access number. Any facsimile equipments calling that number will have their message received by the UA. The UA assumes responsibility for delivering the message to the recipient.

In another short-term mechanism, the service could be made visible to the facsimile user. The UA would act as a gateway. Access to the UA could be done using normal facsimile procedures (stage 1 of 2). The user could then send the UA a "special" facsimile message (stage 2 of 2). The message could consist of two parts: a header and a message body. The header contains delivery instructions. The message body is what is to be delivered. This approach has a major advantage over the first short-term mechanism. Most of a service's capabilities can be taken advantage of (like multiaddressing, deferred delivery, etc.).

By using human-readable graphic characters (e.g., Recommendation T.61), the header could be easily constructed using simple office equipments (typewriters, for instance). Upon receipt, the UA would decompress the header using facsimile

techniques. Then, the UA could interpret the instructions using OCR techniques. Using OCR, however, requires considerable sophistication on the part of the UA.

1.2.2 Long-Term Mechanisms

Long-term mechanisms could allow facsimile equipments to elicit message transmittal information from the operator and electronically transmit it to the UA. At least three reasons suggest this approach: 1) to validate the information, 2) to reduce the chance of transmittal errors, and 3) to reduce UA complexity. Mechanisms meeting these criteria are character transmissions and binary encoding. Binary encoding is similar to character transmissions. They differ mainly in the number of representative bits. Binary encoding uses mostly single binary bits to convey instructions and information. Character transmissions carry the instructions and information in one or more octets.

Sending characters has several advantages:

1. Efficient transmission of complex instructions and detailed information.
2. Characters are communicable over almost any network.
3. Requires no special sophistication within the UA.
4. Can provide a path between the short-term and long-term mechanisms

1.2.3 Transitioning From Short-Term to Long-Term Mechanisms

During the transition from short-term to long-term mechanisms, UAs may have to support both mechanisms. To ease this transition, some commonality between the two is probably desirable. Commonality may be possible if OCR provides the short-term mechanism and character transmissions provides the long-term mechanism. The OCR instruction set could be a subset of the character transmission instruction set. If it were, transitioning from the short-term

mechanism to the long-term mechanism might be as simple as bypassing the UA's character recognition module (See Figure 1-1).
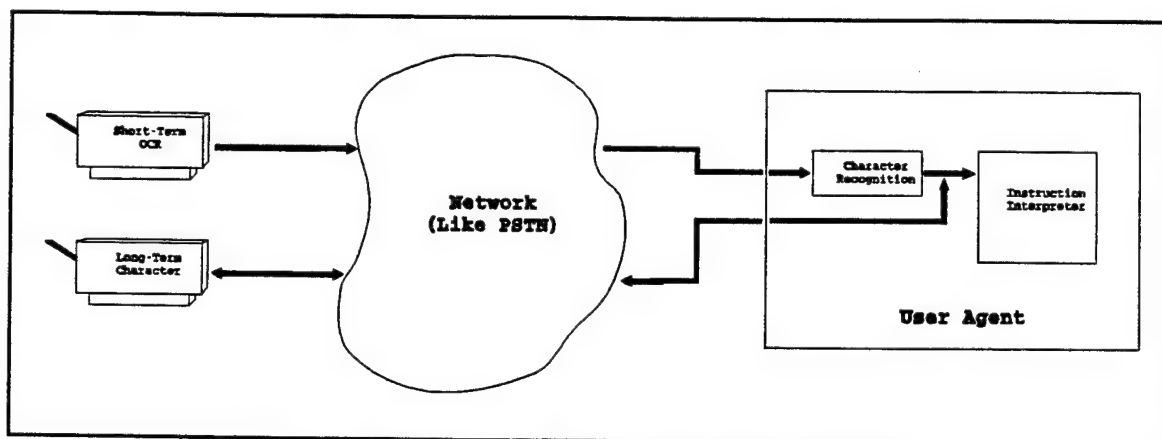
Figure 1-1. UA having OCR and Character Capabilities

If easing the transition from short-term to long-term mechanisms is not a concern, binary encoding may be preferred as the long-term mechanism. It is usually more efficient than character transmissions, and usually compacts more information into fewer bits. A comparison of these two mechanism is shown in Table 1-1.

1.2.4 Facsimile to Text Conversion

Converting facsimiles to text is becoming desirable, especially for personal computers (PCs). Text usually requires less storage space, and less time to print. Plus, text can be easily edited and incorporated into other documents. Automatically transforming text-based facsimiles to character documents requires an OCR capability. This capability could be provided by a UA. In some cases this conversion might be done for text-only terminals. Facsimiles containing renditions of imagery, like line graphics, half-tones, and gray-scales, pose special challenges for OCR. This is especially true if the recipient has a text-only terminal.

Table 1-1. Comparison of Long-Term Mechanisms

Capability	Binary Encoding	Character
Network Independent	yes	yes
Efficient coding of instructions	yes	yes
Requires modifying terminals	yes	yes
Easy implementation	no	no
Requires complex UA	no	no
Provides extensive MHS capabilities	yes	yes
Transmitted instructions verifiable	yes	yes
User friendly (i.e., easy to use)	yes	yes
Likelihood of instruction misinterpretations	very low	low
Usable by Group 3 and Group 4	yes	yes
Compatibility with one or more short-term mechanisms	low	high

1.3 Objectives

There were four main objectives for this study:

1. To assess the effectiveness, reliability, and speed of OCR as a mechanism for interpreting store-and-forward instructions from a facsimile terminal.
2. To assess the effectiveness, reliability and speed of OCR as a mechanism for converting facsimiles to character-based documents. This includes identifying graphics only areas and simulating the graphics with characters.
3. To assess the performance differences between the character and binary encoded methods for transporting

instruction streams from facsimile terminals to store-and-forward systems, and vice versa.

4.	To assess the various methods being considered for providing information to networks (e.g., that use Digital Circuit Multiplying Equipment) on the type of modulation method being used.

2.0 OPTICAL CHARACTER RECOGNITION

Text comes in many forms, both machine-made and handwritten. There are hundreds of type fonts and thousands of print fonts in the world, and each has its own distinctive style and peculiarities.[1] They include serifs, shapes, curvatures, sizes, pitch, line thickness, and so forth. Variations in handwritten characters are even greater. Each person has his own way and style of writing and samples from the same hand are seldom identical in shape or size. The most confusing character pairs are 6/G, D/O, I/1, S/5, 2/Z, and U/V. Mainly because they have very similar topological structures.

A typical OCR system is shown in Figure 2-1. At the input end, the OCR locates the regions where data has been printed or written and segments them into character images. After segmentation, a preprocessor then eliminates random
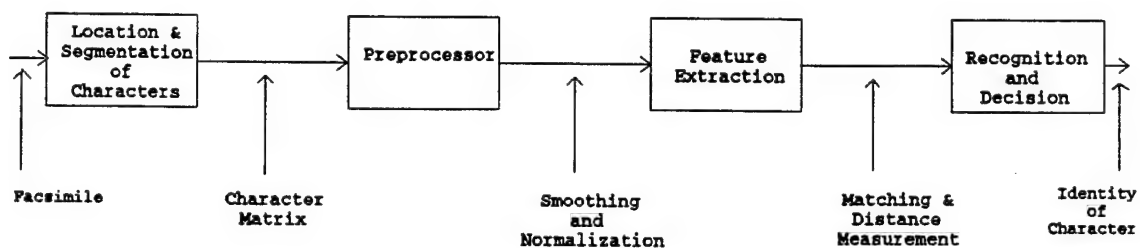
Figure 2-1. Example of an OCR System

noise, voids, bumps, and other spurious components of the segmented characters, if present, and thins the characters. This process is known as smoothing. Sometimes, normalization in size, orientation, position, and other operations are done to help the following stage extract distinctive features. Normalization produces patterns of uniform size or linewidth, fixed boundaries along certain edges (top-left justification), or a preferred orientation (vertical). Doing so usually simplifies feature extraction and improves the recognition rate. After the image is smoothed and normalized, the feature extraction stage extracts the features that allow the system to discriminate correctly one class of characters from others. After the features are extracted, the recognition and decision stage classifies them by comparing them to a list of references and knowledge base. This stage also uses context, distance measurements, shape derivation, shape matching, and

hierarchical feature matching in the form of decision trees. The decision stage is strongly influenced by the extracted features, and a successful OCR is built on the joint operations and performances of the feature detector and the classifier.

There are many techniques for recognizing text. Some include error correction techniques that consider how errors occur and how characters are used in context. Of particular concern is the handling of many different fonts (multifonts). OCR products incorporating some of these techniques are offered by several commercial companies. Their products are quickly becoming more proficient at identifying characters while reducing character misrecognitions.

2.1 Recognition Techniques

Recognition techniques can usually be separated into two classes, serial processing and parallel processing.

2.1.1 Serial Processing

One serial processing technique isolates the primitive characteristics of a character, segments and angles. This approach permits economical description and avoids the necessity for large numbers of template-like tests for each character. By specifying essential parts and their inter-relationships, feature analysis is usually geometrically invariant. It takes advantage of the natural properties of characters. (e.g., the differences between different characters are usually more significant than differences between different renditions of the same character) It is usually insensitive to character size, position, rotation, and to some extent style. Recognition is accomplished by comparing a set of detected features to a stored set of features for each character. A unique fit or one with maximum correlation usually yields the desired character.

Another serial technique uses curve tracing. In this approach the line structure of a character is traced. Then the tracing is matched against a set of

stored tracings to find the best fit. Curve tracing can be sensitive to breaks in lines and problems that arise at nodes containing two or more intersecting lines.

2.1.2 Parallel Processing

Parallel processing usually uses direct optical matching. As a result, it often requires more stringent constraints on the input material. One parallel process is template matching. It is one of the oldest character recognition techniques. With template matching, the degree of a match between an input pattern and each of a stored set of reference patterns is determined. The input pattern is successively fitted to the reference pattern masks or templates. The best fit usually determines the output character. This type of matching is very sensitive to variations in character size, rotation, and style. To overcome these difficulties, either very large numbers of masks are typically provided or elaborate provisions for rotation, translation, and varying magnification are required.

A variant of template matching is peephole matching or n-tuple correlation. N-tuple correlation resembles feature extraction to some degree. By using selected subareas of the image field for template matching, it can distinguish characters by their features. For n-tuple correlation, small-area masks are assigned independent weights. The masks are chosen to effectively separate characters. In some cases, the extent, position, and weight of each area are determined by computer analysis of large samples of input material. In these cases, the n-tuple mask positions are likely to coincide with the characters' features. Particular sets of n-tuple masks define particular characters. Character identification depends on the shape of a given character forbidding certain n-tuple states while others are highly likely.

Coordinate matching is another variant of template matching. With it, the character field is quantized onto a grid. The coordinates of the grid form the basis for representation and matching. The quantized image is analyzed point by point for a binary representation of occupied coordinate points. That representation is then compared with stored quantized templates.

2.2 Multifonts

The recognition of a large number of fonts poses several challenges:

1. Variations in character shapes
2. For some fonts, the lack of distinction between "oh" and "zero" or "one" and "lower case el."
3. Variations in sizes, 10 pitch characters are usually bigger in both width and height than those in 12 pitch and 15 pitch.
4. Variations in pitch, for example, 10, 12, 15 pitch, which corresponds to 10, 12, and 15 characters/in., and proportional spacing. These variations affect the location and segmentation of characters in a typewritten or printed text.
5. Ornaments and serifs of the characters, for example the difference between sans serif fonts like Gothic and Orator and serif fonts like Courier and Elite.
6. Variations in line thickness, for example bold fonts are thicker than regular fonts, while sans serif characters have more uniform widths than those with serifs.
7. Italics whose characters are all tilted to the right
8. Script, which has a cursive type style that simulates handwriting
9. Some characters that may touch each other, for example, the wide characters m and w may touch their neighbors.

2.3 Character Errors

Text is usually garbled by one or more errors:

1. Typographical errors committed by manual keying.
2. Spelling errors committed during text creation.
3. Errors committed by the OCR process.

2 - 4

4. Grammatical errors, semantic errors, consistency errors, and usage errors.
5. Storage and transmission errors.

Typographical errors are usually inconsistent but perhaps more predictable. They could be related to the position of keys on the keyboard and probably result from errors in finger movements. Typographical errors are predominantly of four types: insertion, deletion, substitution, and transposition (reversal). Insertion errors include an additional letter in a string of text (e.g., texxt - extra x). Deletion errors leave out a letter in a string of text (e.g., tet - missing x). Substitution errors means one character replaces another in a string (e.g., txxt - x for e). Transposition (reversal) errors means two characters of a string are reversed. The characters are usually adjacent (e.g., txet - xe for ex).

English is not a phonetic language. There is no direct correspondence between the sound and spelling of a word. Some words have been borrowed from other languages with different spelling and phonetic rules. English also has multiple prefix and suffix forms that serve the same purpose. These typically have only minor variations, e.g., the prefixes *en* and *in* and the suffixes *able* and *ible*. The difference between how a word sounds and how it is actually spelled can result in consistent misspellings. Especially if an author is ignorant of the correct spelling.

OCR interpretation usually introduces substitution and rejection errors. The substitution errors are often limited to the replacement of a character by another whose shape is similar (e.g., u for v). A rejection error indicates that the OCR process was unable to recognize a character confidently. For rejection errors a place holder for the unknown character is often placed in the text (e.g., a "?").

Words can be classified according to their word class (noun, verb, adjective, article, etc.). Then, the structure of sentences can be checked to ensure they are properly constructed and syntactically correct. Once syntax is checked, grammar can also be checked. This includes proper word use and punctuation. The use of different correct spellings for a given word, such as *grey* and *gray*, are consistency

errors. This includes the use of an appropriate case (upper or lower) for proper names (e.g., IBM, Jim).

Storage and transmission errors relate to specific encoding and transmission mechanisms. For example, during a facsimile transmission, an error could cause the previous scan line to be repeated one or more times. This could distort characters, possibly rendering them unrecognizable.

2.4 Context Sensitivity

Misinterpreted or unidentified characters can be reduced by analyzing their use in context. Some characters can be identified (or eliminated) by using contextual knowledge:

1. Probabilities of letters, letter pairs, letter triples, etc.
2. Probabilities of words.
3. Legal combinations of letter pairs, triples, etc.
4. A lexicon, or list of words acceptable in global or local context.
5. A grammar describing the syntax of the language .

Contextual knowledge is used in one of two ways usually. The first is in conjunction with shape knowledge when classifying characters. The second is as a post processing step after characters have been classified. Both help reduce the number of unidentified and misinterpreted characters.

2.5 Speed and Reliability

Today, several vendors make OCR products.[2] (See Table 2-1.) Their products are able to recognize multiple font text (OmniFont) at rates up to 1,866 characters/sec on IBM PCs or compatibles. Typically, they have an accuracy greater than 99 percent, if paper skew is minimized. Accuracy is also dependent on the quality of the original. To improve accuracy some of these products use

contextual knowledge. Several are able to recognize facsimiles. They usually prefer that facsimiles be sent at the highest possible resolution. In general, these products are designed to work with 300 dpi and can accurately recognize characters larger than 6 point. For characters with a size of 6 point or smaller, higher resolutions are usually needed.

Table 2-1. OCR Vendors

Manufacturer	Omni Fonts	Font Sizes (Point)	Proportional Spacing	Fax	Skew (degrees)	Recognition Rate (chars/sec)	Accuracy
Caere (OmniPage)	✓	6 to 72	Yes	Yes	-	1,866	99.77
CTA	✓	6 to 72	Yes	Yes	-	1,333[1]	100.
Ocron (Perceive)	✓	8 to 36	Yes	-	-	770	99.61
OCR Systems (ReadRight)	✓	6 to 72	Yes	-	-	581	99.15
Recognita (Recognita Plus)	✓	6 to 24	Yes	-	-	1376	99.60
ExperVision (TypeReader)	✓	6 to 64	Yes	-	-	858	99.46
Calera (WordScan Plus)	✓	6 to 28	Yes	Yes	± 2	735	99.30

[1] Uses an OCR accelerator board

3.0 INSTRUCTION STREAM RECOGNITION

OCR could be used for automatic instruction stream recognition. In one approach, a sender would compose a cover sheet that contains the delivery instructions. He might use a typewriter, for instance. Then, the sender would fax the cover sheet and message to the delivery service. The delivery service's UA would use OCR to interpret the instructions on the cover sheet. After deciphering the instructions, the UA would assume responsibility for delivering the message.

In general, given the speed, accuracy, and font and pitch capabilities of commercial products, instruction recognition is likely to approach 100% accuracy. Commercial OCR products can recognize a large variety of fonts, including those transmitted by facsimile. Accuracy depends mostly on image quality and transmission resolution. Accuracy can be improved by using contextual knowledge. Ascertaining how well some current OCR products recognize low quality cover sheets could be addressed in future studies.

Although commercial OCR products are approaching 100% accuracy, given the large number of characters processed, some errors are likely to occur (e.g., with an accuracy of 99.9%, a page with 2000 (80 x 50) characters is likely to have 2 errors). Additional errors from missing, unknown, or erroneous instructions can also occur. To resolve errors, an error recovery mechanism should be available. It should account for unidentified or misinterpreted characters and missing, unknown or erroneous instructions.

In one recovery mechanism, the UA would reject any message where the sender's instructions are unclear. The sender could be notified of the rejection via a report faxed to the sender's terminal. The rejection notice would detail why the transmission failed. For example, no recipient address was given, or unidentifiable characters were detected in the address. After making appropriate corrections, the sender could resubmit his message.

In another recovery mechanism, the UA would store the message and notify the sender that the delivery instructions are unclear. Like the rejection notice, this notice might be faxed to the sender. It would detail why the instructions are

unclear. Plus, it might explain how to correct the instructions. For example, it might request a submission of the corrected cover sheet. Or, it might take advantage of Dual-Tone Multi-Frequencies (DTMF). In the latter case, the sender might be instructed to telephonically call the UA. The UA, when called, could verbally prompt the caller for corrections. The caller could indicate what the corrections are using DTMF. For example, if the UA believes a character might be either a "u" or a "v," it might ask the sender to press either a "1" or "2" or a "#" if neither. To simplify and streamline the process, the errors on the notification could be numbered, and allowed DTMF responses for each could be supplied. Then, during the telephone conversation, the UA need mention just the error number (e.g., UA says "error 1"). The user would then be responsible for supplying an appropriate DTMF response. After all errors are corrected, the UA could then deliver the stored message.

4.0 CONVERTING FACSIMILES TO TEXT DOCUMENTS

Almost any type of document can be sent via facsimile. Typical transmissions carry handwritten notes, typed business letters, magazine pages, or color photos. Converting these facsimiles to text could require separating text, handwriting and imagery, and performing handwriting recognition, text recognition and image processing. The mix chosen depends on the document and the receiving terminal's capabilities. Some PCs and PC-based word processors permit both text and imagery in a single document (e.g., WordPerfect). As fax terminals, the PCs could convert facsimiles into forms usable by their software packages. These packages could then store, display, print, or retransmit the converted facsimile. The conversion process might consist of separating text, handwriting, and imagery, and performing text recognition, handwriting recognition, and image processing. Text only terminals might require the same basic processing steps. Images, however, might have to be represented using text characters.

4.1 Separating Text, Handwriting, and Images

The separation of text, handwriting and imagery is usually done before recognition techniques are applied. Separating the three can be done without recognizing individual characters, regardless of string orientation and font size or style.[3] One method uses simple heuristics based on the characteristics of text strings. With this method, the separation process is broken into five steps:

- Connected component generation,
- Area/ratio filter,
- Collinear component grouping,
- Logical grouping of strings into words and phrases,
- Text string separation.

The connected component generation involves grouping eight connected black pixels (assuming a black image on white background). The eight connected pixels belonging to individual characters or graphics are enclosed in circumscribing rectangles. Each rectangle identifies a single connected component. (See

Figure 4-1 and Figure 4-2.) The output from the connected component generation process is an array that specifies the maximum and minimum coordinates of the circumscribing rectangles of connected components, the coordinates of the top and bottom seeds of each connected component, and the number of black pixels. Each connected component is either rejected or accepted as a member of a text string based on its attributes (size, black pixel density, ratio of dimensions, area, position within the image, etc.).

Figure 4-1. Test Image

An initial examination of connected component attributes (Area/ratio filter) can reduce the working set of connected components to one that contains a higher percentage of characters. In general, a mixed text/graphics image produces connected components of widely varying areas. The larger connected components usually represent the larger graphic components of the image. By obtaining a histogram of the relative frequency of occurrence

Figure 4-2. Rectangles Enclosing Connected Components

of components as a function of their area, it is possible to set an area threshold that broadly separates larger graphics from text components. A similar filtering

can be done with connected component dimensional ratio attributes. Very long lines are unlikely to be text characters.

The collinear component grouping step logically connects characters into strings that lie along any given straight line. Further grouping may take place by examining the distance between characters (logical grouping of strings into words and phrases step). By comparing the intercharacter distance with the interword gap and intercharacter gap thresholds, the string can be segmented into logical character groups (words or phrases). Only if components belong to a logical character group can they be considered a members of a valid text character string.

In the text string separation step, text is physically separated from graphics. Two images can be made. One contains only text; the other contains only graphics. This involves moving all connected components corresponding to strings from the graphics image to the text image. In the graphics image, this also involves replacing black pixels, belonging to marked connected components, with white pixels. It is important that only those black pixels that originally formed a particular connected component should be moved, not all black pixels within the area of the circumscribing rectangle. By using the black pixel seeds, only the pixels that originally formed a particular connected component are moved. Once text and graphics are separated, recognition techniques can be applied.

4.2 Text Recognition

Generic text recognition requires recognizing multifonts and variable size characters while removing tilt restrictions.[4] Template matching systems are often ill-suited to this task. They tend to be font sensitive. Feature-based systems, on the other hand, tend to be font-insensitive. Recognition errors can be reduced by taking advantage of statistics on character errors and incorporating context sensitivity. (See Section 2.0, "Optical Character Recognition.")

4.3 Handwriting Recognition

Variations in handwritten characters are greater than those in type fonts. Each person has his own ways and styles of writing. Character samples written by the same hand are never identical in shape or size. There are an infinite number of possible character shapes. The large variability in handwriting may be attributed to writing habits, style and care in writing, education, region of origin, mood, health and other conditions of the writer. Writing instruments and writing surfaces also play a role.

Handwriting recognition is in its infancy. To improve recognition handwriting is often constrained. Constrained handwriting requires the writer to print characters carefull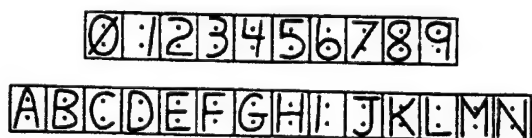y in certain areas. (See Figure 4-3.) Emphasis is on accuracy of machine reading rather than on speed and flexibility of writing. When done properly, 99% accuracy is achievable. Compare this to 96% for most humans when reading handprinting in absence of context.[5] Characters with similar topological structures are the most confusing pairs. Poor penmanship often makes recognition even more difficult. Examples of confusing pairs are 6/G, D/O, I/1, and U/V.

Ø123456789
ABCDEFGHIJKLMN

Figure 4-3. Constrained Handwriting

Unconstrained handwriting is even more difficult.[6] Recognition is harder because characters can run together. There are three basic types of handwriting, discrete printing, run together discrete printing, and cursive writing. (See Figure 4-4.) For all three, prior to recognition characters are usually separated. Discrete characters are usually the easiest to separate. They have spaces between characters. Run together discrete characters are easier to separate than cursive characters. Discrete characters consist of one or more strokes. Cursive

Spaced Discrete

Run-on discrete

Cursive

Figure 4-4. Types of Handwriting

characters often consist of a single stroke and more than one character can be made with that stroke. Since discrete characters consist of one or more strokes, character separation is usually done after a stroke. For cursive writing, separation is usually done within strokes.
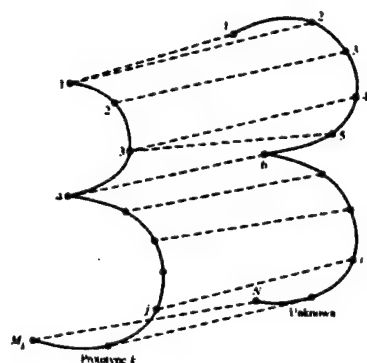

Figure 4-5. Elastic Matching

A popular technique for recognizing both discrete and cursive handwriting is elastic matching. In elastic matching the character to recognize is compared to a prototype. (See Figure 4-5.) Explicit letter separation is not performed. Rather, elastic matching evaluates all possible separations and simultaneously obtains the best combination of segmentation and recognition. Elastic matching is insensitive to minor perturbations of input letter shapes relative to prototype letter shapes.

4.4 Imagery

Today, most fax transmissions are bi-level (black and white). As a result, gray scales and color are usually severely distorted. Some sense of an original's tonal range can be restored by applying half-toning or dithering techniques during the facsimile process. In conventional bi-level systems the scanning threshold is normally fixed midway between peak black and peak white. So any gray scale values near the threshold are drastically altered in the output image. (See Figure 4-6.) To reduce these distortions, the threshold can be varied from pel to pel. Over a number of neighboring pels the visually perceived value approximates the average gray scale values of those pels. (See Figure 4-7.)
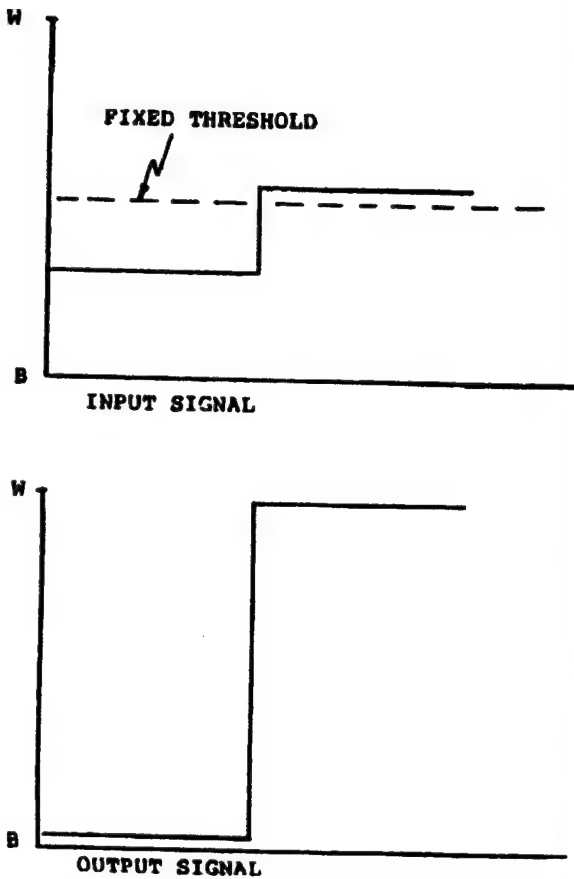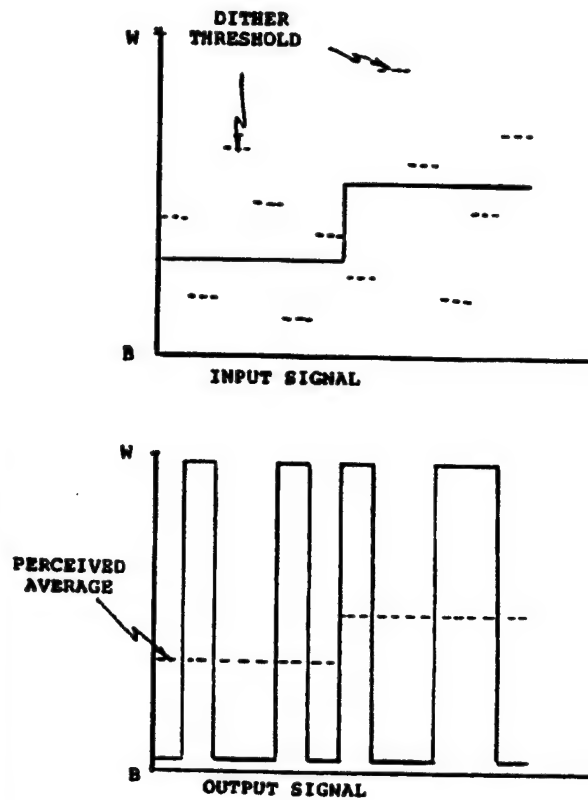
Figure 4-6. Fixed Level Scan

Figure 4-7. Varied Threshold

Images displayed on text only terminals can use characters to approximate perceived gray scale values. Character positions within an image can be regarded as consisting of a two-dimensional cell array.[7],[8] Printed characters (or combination of overprinted characters) for each cell can be chosen according to the cell's average print density. (See Table 4-1.) By doing so a pictorial representation may be generated. (Compare Figure 4-8 to Figure 4-9.) These images are likely to be inferior, however, given the coarseness of the micropattern. Plus, as image size approaches cell size, images often become unrecognizable.

Nevertheless, images displayed using characters as the micropattern can, in some cases, give a recipient a sense of the original.

Table 4-1. Character Density Codes Examples

Overprinted Character Combinations	Estimated Density Values
blank	0.0
-	0.15
=	0.22
+	0.25
)	0.29
1	0.33
Z	0.37
X	0.40
A	0.42
M	0.45
O-	0.53
O=	0.56
O+	0.60
O+,	0.64
O+,.	0.67
O+,.=	0.79
OX'.-	0.85
OX'.HC	0.89
OX'.HB	0.93
OX'.HBV	0.97
OX'.HBVA	1.00

Figure 4-8. Half-tone of Original

Figure 4-9. Gray Scale Image Using Characters

5.0 COMPARISON OF CHARACTER AND BINARY ENCODING METHODS

Whether to use character or binary instruction streams between facsimile terminals and store-and-forward systems partially depends on how the instructions are conveyed. In one short-term mechanism, instructions are conveyed on a cover sheet using Group 3 facsimile procedures. The store-and-forward system uses OCR to "read" the instructions. In the long-term, the general philosophy of this approach could be retained. That is, using a cover sheet to convey instructions. By taking advantage of Group 3's new character mode, the cover sheet could consist of characters. In another approach for conveying instructions, no cover sheet would be used. Instead, store-and-forward instructions would be embedded within the "bit-oriented" Group 3 protocol. One advantage of this approach is that capabilities can be negotiated and instructions can be easily verified.

Using a character mode cover sheet has several advantages:

- Instructions can be self-contained,
- Instructions can be easy to add,
- No modification of the Group 3 protocol is necessary,
- Compatible with OCR'd fax cover sheet,
- Store-and-forward User Agent modifications are minimized.

Character-based instructions can encapsulate both the instruction and its associated data. For example, a person's name could be conveyed via the following instruction:

NM,"Mr. John Doe"

"NM" could be the mnemonic indicating that a person's name follows.

Given such an instruction structure, instructions can be easily added to the instruction repertoire simply by giving new instructions unique mnemonics. Since Group 3's character mode option is used, no Group 3 protocol modifications are

necessary. In addition, modifications to Store-and-forward user agents are reduced to adding new instruction processing.

Using character mode cover sheets is a natural extension of the fax cover sheets that store-and-forward systems might read for instructions. The store-and-forward can process the instructions like it does with a fax cover sheet; except OCRing the cover sheet is unnecessary.

A character mode cover sheet has a couple disadvantages:

- Capabilities negotiations may be impractical,
- Faulty instructions could be difficult to correct.

With the cover sheet approach a secondary mechanism may be needed to negotiate capabilities or correct faulty instructions. (See Section 3.0.)

Embedding store-and-forward instructions within the Group 3 protocol has several advantages.

- Capabilities may be negotiated,
- Commands are encapsulated,
- Commands are verifiable,
- No cover sheet is needed.

Embedding instructions within the Group 3 protocol makes capability negotiation possible and makes verifying commands possible. For example, bits signifying store-and-forward capabilities can be added to DIS/DCS. Plus, an instruction verification mechanism can be added to the protocol. For example, if an illegal or ill-formed instruction is sent, the store-and-forward system can immediately alert the fax terminal. Modifying the protocol eliminates the need for a cover sheet, shortening transmission times.

Embedding instructions within the Group 3 protocol also has disadvantages:

- Protocol changes are necessary
- Excludes installed base of fax equipments

Embedding instructions within the Group 3 protocol is incompatible with the installed base of fax equipments. A mechanism permitting these equipments to communicate with store-and-forward systems may also be necessary.

Table 5-1 compares the character and bit encoded instruction streams.

Table 5-1. Comparison of Character and Bit Encoded Instruction Streams

Capability	Character	Bit
Requires Group 3 Protocol Modifications	No	Yes
Requires Store-and-Forward Modifications	Yes	Yes
Verification of Transmitted Instructions	Poor	Good
Compatibility with Short-term Mechanisms	High	Low
Suitability for "Cover Sheet" Mechanism	High	Low
Suitability as Group 3 Protocol Modification	Low	High

6.0 FACSIMILE TRAFFIC COMPRESSION OVER LONG DISTANCES

High speed facsimile modems and proprietary enhanced services challenge the ability of long distance equipments to efficiently transport them. At present, the CCITT is working on V.FAST, a high-speed modem. V.FAST defines a voice and data modem that reaches the maximum theoretical bit rate (26 Kb/s) of the analog telephone system. Proprietary enhanced services are often added by a manufacturer via the Non-Standard Facilities (NSF).

Facsimiles are often ported transcontinentally via Digital Circuit Multiplying Equipment (DCME) and Packetized Circuit Multiplying Equipment (PCME). These equipments often demodulate and remodulate facsimile traffic to provide efficient transport and to reduce the introduction of errors. In general, DCMEs and PCMEs must 1) identify facsimile and modem traffic, and 2) identify the modulation scheme used if the traffic is to be demodulated and remodulated. In addition, they should accommodate switches between voice, facsimile, and data within the same call. Furthermore, automatic call routing devices that are used at installations where several terminals are operated (e.g., telephone, modem, facsimile equipment) should be considered.

DCMEs reduce the cost of long distance transmissions by concentrating a number of input channels (trunk channels) onto a smaller number of output channels (bearer channels). This is done by connecting a trunk channel to a bearer channel only for the period that the trunk channel is active. That is, when the channel is carrying a burst of speech or voice-band data (e.g., facsimile). For average conversations, one direction of transmission is usually active for 30 to 40 percent of the time. When the number of trunks is large, the statistics of speech and silence distributions permit a significantly smaller number of bearer channels to be used.

On transcontinental links facsimile traffic often becomes quite heavy. To efficiently transport facsimiles and to provide greater bandwidth for voice traffic, facsimiles can be broken into low-speed data and high-speed data. The high-speed data might carry the T.4 coded image information. The low-speed data might carry information like the modem control information and might be treated as voice

traffic. If it is, quiet periods can be taken advantage of. Care must be taken however to maintain signal timings. Failure to do so could interfere with the proper operation of the facsimile equipment at either end.

While voice traffic and low-speed data traffic (below 9600 b/s) are virtually unaffected by DCMEs, higher-speed data traffic can be affected. Problems introduced by DCMEs (e.g., bit errors) rise in proportion to the data signalling rate of the modem. As a result, DCMEs need to identify high-speed facsimile and modem traffic (above 9600 b/s) so the facsimile and modem traffic can receive special treatment.

For some high-speed facsimile and modem traffic, the special treatment consists of the DCMEs (and PCMEs) demodulating the modem signal at their input and remodulating it at the receiving DCME's output. When this is done specific information like the modulation method employed must be determined. For standardized facsimile features this task is usually straightforward. Proprietary modulation methods, however, are more difficult to determine. Proprietary methods can be invoked via Group 3's nonstandard facilities mechanism.

Another approach is to send the facsimile and modem traffic "as is" using 40 K b/s lines. No demodulation or remodulation is done. Although inefficient, this approach is often taken if the modulation method is undeterminable.

Several mechanisms have been proposed to help identify facsimile traffic and the modulation method employed. Most of these are terminal based solutions and include

- the modification of the Calling Tone Signal (CNG),
- transmission of the Digital Command Signal (DCS) together with the Non-Standard Facilities Set-Up Signal (NSS)
- Dual-tone MultiFrequency (DTMF) tones, and
- a new signal specifically for network use.

All of these approaches, except where noted, fail to address existing equipments.

CNG is usually sent by automatic calling facsimile units and is optional for manual units. In practice, however, some automatic calling units do not send CNG and neither do most manual units. CNG detection by DCMEs could be used to identify facsimile calls. To provide consistent identification, CNG would have to be mandatory for both automatic and manual units. It does not address which modulation methods will be used, however. Identifying the modulation method could be accomplished by amending the CNG in at least two ways. In the first, DTMF tones could be used. In the second, modulation information could be transmitted in an HDLC frame structure using the V.21 modulation system. Using V.21 as a common signalling rate would make the signal easily detectable by DCMEs. Nevertheless, this approach does extend the handshake sequence, which many users feel is already too long.

Sending the DCS with the NSS permits identification of all standardized features of the non-standard facsimile call. As a result, DCMEs could be informed of the modulation scheme via the Facsimile Information Field (FIF) of the DCS frame. This approach helps identify the modulation methods used but does not help DCMEs identify facsimile traffic.

Several techniques have been proposed using DTMF tones. In one approach, the originating facsimile sends a DTMF during call establishment. The tone could be used by both DCMEs and the user device. The tone, however, would have to be on for a relatively long period to assure that both the network device and the user device have adequate time to detect it. In addition, generating this signal using a telephone during manual origination may be impractical, for the same reason.

In another approach, dialing sequences are used. It requires a modification of the international dialing plan. In this approach a unique code (e.g., the "#" or the "*") would be inserted between the international access code and the country code on international calls to indicate that a facsimile or data call was being established. This would provide a positive identification to the DCMEs, and it could be expanded to include the modulation scheme being employed. It could be easily accommodated by either automatic or manual call originators, and it does not require DCMEs to detect new in-band signals. The type of call would be

directly communicated from the switching equipment to the DCMEs based on the dialed sequence. The major benefits are that it addresses the installed base and manually originated calls. The disadvantages are that it requires a change to the international dialing plan. Such a change may be impractical.

Of the approaches discussed, none appears adequate for both existing and future equipments. One possible solution might be to use only the best attributes of each. For example, making CNG mandatory for facsimile and data modems and amending it to indicate the modulation method and type of call would allow DCMEs to easily detect new facsimile equipments and data modems. In addition, automatic call routing devices (ACRDs) could distinguish between voice, facsimile, and modem calls. Modifying the international dialing plan would permit the identification of existing equipments by DCMEs. For ACRDs, existing equipments might be handled as follows

1) A voice call is assumed if the originating equipment issues no calling tone,
2) A facsimile call is assumed if the calling tone is 1100 Hz,
3) Any other calling tone is assumed to be a data modem call, and
4) Equipments issuing no calling tone must be manually switched.

Mandatory CNG has additional support in that it will very likely be proposed at the next CCITT Study Group VIII meeting, and in that the CCITT plans to use CNG for automatic terminal selection. Recommendation T.30 is being modified to permit automatic terminal selection between facsimile equipments, telephone answering machines, and telephone answering and recording machines. This automatic terminal selection relies on CNG detection for early and reliable identification of incoming facsimile calls. There is some disagreement, however, that automatic terminal selection can really work. For example, the value of one timer is extremely critical, and some feel that it is not possible to give it a value whereby all equipments are accommodated.

7.0 SUMMARY AND RECOMMENDATIONS

OCR could be used as a mechanism for interpreting store-and-forward instructions from a facsimile terminal and could be used for converting facsimiles to text. It is fast and is able to interpret most fonts and font sizes. Its accuracy is approaching 100 percent. Nevertheless, its accuracy is low enough that misinterpretations can occur. As a result, a mechanism for correcting misinterpretations is probably necessary. At least two mechanisms are possible. In the first, the store-and-forward sends a fax containing the suspect instructions or text back to the sending facsimile equipment. In the second, the sender might use DTMF to correct errors.

OCR can also be used to identify a page containing imagery. Text characters could then be used to "draw" the imagery. The resulting images, however, are likely to be very coarse and might give only a sense of the original.

Character or binary encoded instruction streams can be used with instruction "cover sheets" or a modified Group 3 protocol, respectively. The former is compatible with UAs using OCR to "read" a fax-encoded (e.g., T.4) instruction cover sheet. An error correction mechanism might be needed for this approach, however. A protocol modification using a binary encoded instruction stream needs no cover sheet and can incorporate an error correction mechanism. In general, the binary encoded instruction stream is incompatible with short-term mechanisms, however.

Both mechanisms could be combined. The Group 3 protocol could be modified to include both as options: a store-and-forward instruction mode (binary encoded instruction stream), and a store-and-forward cover sheet (character instruction stream). These could be mutually exclusive options. The store-and-forward instruction mode would incorporate store-and-forward instructions into the Group 3 protocol. The store-and-forward cover sheet would use Group 3's character mode option to carry store-and-forward instructions to a UA. If neither is chosen, a UA could assume that the cover sheet is bit-mapped and must be "read" using OCR techniques (standard mode). Combining these options allows existing

facsimile terminals to send store-and-forward messages and provides an evolutionary path for future facsimile terminals.

For long distance equipments (e.g., DCMEs and PCMEs) and automatic call routing devices, identifying facsimile and modem traffic could be made easier if two approaches are used: 1) making CNG mandatory for new facsimile and data modems and amending CNG to indicate the modulation method and type of call, and 2) modifying the international dialing plan to permit the identification of existing equipments.

REFERENCES

[1]   Tzay Y. Young, King-Sun Fu, "*Handbook of Pattern Recognition and Image Processing*," pp. 569-586, Academic Press, Inc., 1986

[2].   Mitt Jones, "OCR Gets You From There to Here," pp. 267-319, PC Magazine, July 1992, Volume 11, Number 13.

[3]   L.A. Fletcher, R. Kasturi, " A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images," *IEEE Trans. Pattern Anal. and Machine Intel.*, vol. 10, No. 6, November 1988

[4]   S. Kahan, T. Pavlidis, H. Baird, "On the Recognition of Printed Characters of Any Font and Size," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no.2, 1987

[5]   C.Y. Suen, M. Berthod, and S. Mori, Automatic recognition of handprinted characters - the state of the art, *Proc. IEEE 68*, 469-487 (April 1980)

[6]   C.C. Tappert, C.Y. Suen, T. Wakahara, "The State of the Art in On-Line Handwriting Recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-12, no. 8, 1990

[7]   I.D.G. Macleod, "Pictorial Output with a Line Printer, *"IEEE Trans. Computers,"* February 1970

[8]   K. Knowlton, L. Harmon, "Computer-Produced Grey Scales," *Computer Graphics and Image Processing,"* Academic Press, vol. 1, no. 1, April 1972